
CURRICULUM VITAE

NAME: **Ka Yee Yeung-Rhee** (also known as **Ka Yee Yeung**)

POSITION TITLE: Professor, School of Engineering and Technology, University of Washington Tacoma, WA.
Adjunct Professor, Department of Microbiology, School of Medicine, University of Washington, Seattle, WA.

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
University of Waterloo, Ontario, Canada	B.Math	05/95	Actuarial Science & Computer Science
University of Waterloo, Ontario, Canada	M.Math	08/96	Computer Science
University of Washington, Seattle, WA	M.S.	12/98	Computer Science
University of Washington, Seattle, WA	Ph.D.	12/01	Computer Science

A. Personal Statement

My research focuses on the development of methods and software tools that leverage cloud and container technology that will allow biomedical researchers to build modular reproducible bioinformatics workflows. In particular, I am strongly committed to develop publicly available software tools for the biomedical community to facilitate reproducible research of big biomedical data. My team has extensive experience in the development of cloud-based software tools for bioinformatics applications.

I teach undergraduate and graduate level computer science courses at UW Tacoma, including data structures, algorithms, and bioinformatics. I also have extensive experience mentoring research scientists, post-doctoral fellows, Ph.D., Master's and undergraduate students. In particular, I am strongly committed to enhancing the diversity of the research workforce.

B. Positions and Honors

Positions and Employment

9/1996-12/1997 Teaching Assistant, Department of Computer Science & Engineering, University of Washington, Seattle, WA
1/1998-12/2001 Research Assistant, Department of Computer Science & Engineering, University of Washington, Seattle, WA
1/2002-5/2004 Research Scientist/Senior Fellow, Department of Microbiology, University of Washington, Seattle, WA
6/2004-6/2011 Research Assistant Professor, Department of Microbiology, University of Washington, Seattle, WA
7/2011-6/2014 Research Associate Professor, Department of Microbiology, University of Washington, Seattle, WA
7/2014-8/2019 Associate Professor, School of Engineering and Technology, University of Washington, Tacoma, WA
7/2014-8/2019 Adjunct Associate Professor, Department of Microbiology, University of Washington, Seattle, WA
9/2019-present Professor, School of Engineering and Technology, University of Washington, Tacoma, WA
9/2019-present Adjunct Professor, Department of Microbiology, University of Washington, Seattle, WA

Other Experience and Professional Memberships

2008	NIH-R21 Peer Review Committee: Exploratory/Developmental Research Grant Award, ad hoc reviewer
2009	NIH Peer Review Committee: ARRA Challenge Grants, stage 1 reviewer
2014	NIH-R01 Peer Review Committee: Computational Analyses Exploiting Reference Epigenomic Maps, ad hoc reviewer
2015-2018	Member of the NIH BD2K-LINCS Data Working Group
2016-2018	Member of the NIH Commons Framework Working Group on Workflow Sharing and Docker Registry
2017-2018	Co-Chair in the search for the Executive Vice Chancellor for Academic Affairs at UW Tacoma
2017-2019	Elected academic leadership experience. Serve as the Vice Chair (2017-2018) and Chair (2018-2019) of Faculty Assembly, University of Washington Tacoma
2018-2019	Member of the University of Washington Senate Executive Committee

Honors

1995	Honorable Mention in the Outstanding Female Undergraduate Student Competition, Computing Research Association.
1995	ITRC Scholarship, University of Waterloo, Ontario, Canada.
1995	Provost Scholarship, University of Waterloo, Ontario, Canada.
1995	NSERC Scholarship, National Science and Engineering Research Council, Canada.
1996	Alan George Student Leadership Award, University of Waterloo, Ontario, Canada.
2002	Fast breaking paper in Computer Science, "Validating Clustering for Gene Expression Data", ISI Essential Science Indicators.
2003	Fast moving fronts in Computer Science, "Model-based Clustering and Data Transformations for Gene Expression Data", ISI Essential Science Indicators.
2017	Distinguished Women Award, University of Washington Tacoma
2019	Distinguished Research Award, University of Washington Tacoma

C. Contributions to Science

Complete list of publications in MyNCBI:

<https://www.ncbi.nlm.nih.gov/myncbi/ka%20yee.yeung-rhee.1/bibliography/public/>

1. My early publications focused on *unsupervised machine learning methods* (also known as clustering, or pattern discovery algorithms) for gene expression data. Given the high dimensionality of gene expression data, it is essential to find patterns from the data to facilitate interpretation and analyses. In particular, we directly addressed the question "which clustering algorithm should I use?" since there are many clustering algorithms in the literature and they all yield clusters. These publications provided a systematic framework for comparing clustering algorithms using gene expression data. In addition, we showed the merit of using a probabilistic framework to cluster gene expression data.
 - a. **Yeung K. Y.**, Haynor D. R., Ruzzo W. L. Validating Clustering for Gene Expression Data. *Bioinformatics* 2001, 17:309-318.
 - b. **Yeung K.Y.**, Ruzzo W.L. Principal Component Analysis for Clustering Gene Expression Data. *Bioinformatics* 2001, 17:763-774.
 - c. **Yeung K. Y.**, Fraley, C., Murua A., Raftery A. E., Ruzzo W.L. Model-based Clustering and Data Transformations for Gene Expression Data. *Bioinformatics* 2001, 17:977-987.
 - d. Young W. C., **Yeung K. Y.**, Raftery A. E. Model-based clustering with data correction for removing artifacts in gene expression data. *Annals of Applied Statistics* 2017, 11(4): 1998-2026.
2. I also developed *supervised machine learning* methods for bioinformatics applications. We have built on the Bayesian Model Averaging (BMA) framework. BMA is a multivariate variable selection (or feature selection) method that accounts for dependencies between variables, as well as uncertainty of model selection by averaging over multiple models. We have developed BMA methods to infer gene networks integrating multiple data sources. Gene networks summarize the relationships between biological entities. We formulate network inference as a series of variable selection problems. In other words, we aim to identify candidate regulators (i.e. parent nodes) from high-dimensional biological data for each target gene.

Our network inference methods guide the search of regulators in the model space using prior probabilities derived from multiple data sources.

- a. **Yeung K.Y.**, Dombek K.M., Lo K., Mittler J.E., Zhu J., Schadt E.E., Bumgarner R.E. and Raftery A.E. Construction of regulatory networks using time series microarray data in genotyped yeast segregants. *PNAS* 2011, 108(48): 19436 - 41.
 - b. Young W.C., Raftery A.E., **Yeung K.Y.** Fast Bayesian Inference for Gene Regulatory Networks Using ScanBMA. *BMC Systems Biology* 2014, 8:47.
 - c. Young W.C., Raftery A.E., **Yeung K.Y.** A Posterior Probability Approach for Gene Regulatory Network Inference in Genetic Perturbation Data. *Mathematical Biosciences and Engineering (MBE)* 2016, 13(6): 1241-1251.
 - d. Hung L.H., Shi K., Wu M., Young W.M., Raftery A.E., **Yeung K.Y.** FastBMA: Scalable Network Inference and Transitive Reduction. *Gigascience*. 2017, 6(10):1-10.
 - e. Liang X., Young W.M., Hung L.H., Raftery A.E., **Yeung K.Y.** Integration of multiple data sources for gene network inference using genetic perturbation data. *Journal of Computational Biology* 2019. <https://doi.org/10.1089/cmb.2019.0036>
3. I contributed to biomarker discovery using 'omics' data. Specifically, I developed systematic frameworks integrating various 'omics' data. Towards this end, I collaborated with clinicians, oncologists and contributed to crowd sourcing challenges.
- a. Oehler V.G.*, **Yeung K.Y.***, Choi Y.E., Bumgarner R.E., Raftery A.E., and Radich J.P. The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data. *Blood* 2009: 114, 3292-3298. *Co-first authors.
 - b. **Yeung K.Y.**, Gooley T.A., Zhang A., Raftery A.E., Radich J.P., and Oehler V.G. Predicting relapse prior to transplantation in chronic myeloid leukemia by integrating expert knowledge and expression data. *Bioinformatics* 2012, 28(6): 823-830.
 - c. Fourati S., Talla A., Mahmoudian M., Burkhart J.G., Klen R., Henao R., Aydin Z., **Yeung K.Y.**, Ahsen M.E., Almugbel R., Jahandideh S., Liang X., Nordling T.E.M., Shiga M., Stanescu A., Vogel R., The Respiratory Viral DREAM Challenge Consortium, Pandey G., Chiu C., McClain M.T., Woods C.W., Ginsburg G.S., Elo L.L., Tsalik E.L., Mangravite L.M., Sieberts S.K. A crowdsourced analysis to identify ab initio molecular signatures predictive of susceptibility to viral infection. *Nature Communications* 2018, 9:4418.
4. I also develop containerized software tools that are easy-to-use and easy-to-deploy to enhance reproducibility of bioinformatics analyses. My lab brings the latest technology to biomedical scientists by adding a graphical user interface to container-based tools, featuring gene network inference and RNA-sequencing data analyses as proof-of-concept examples.
- a. Hung H.L.*, Kristiyanto D.*, Lee S.B.*, **Yeung K.Y.** GUIDock: Using Docker containers with a common graphics user interface to address the reproducibility of research. *PLoS One* 2016, 11(4):e0152686. *Co-first authors. Software available at <https://github.com/BioDepot/GUIDock-X11>
 - b. Mittal V., Hung L.H., Keswani J., Kristiyanto D., Lee S.B. and **Yeung K.Y.** GUIDock-VNC: Using a graphical desktop sharing system to provide a browser-based interface for containerized software. *Gigascience* 2017, 6(4): 1-6. Software available at <https://github.com/BioDepot/GUIDock-VNC>
 - c. Hung H.L., Hu J., Meiss T., Ingersoll A., Lloyd W., Kristiyanto D., Xiong Y., Sobie E.A., **Yeung K.Y.** Building containerized workflows using the BioDepot-workflow-Builder (BwB). *Cell Systems*, in press. [bioRxiv 099010](https://doi.org/10.1016/j.celsys.2019.09.010) [Pre-print]. Software available at <https://github.com/BioDepot/BioDepot-workflow-builder>
 - d. Almugbel R., Hung L.H., Hu J., Almutairy A.M., Ortogero N.E., Tamta Y. and **Yeung K.Y.** Reproducible Bioconductor Workflows Using Browser-Based Interactive Notebooks And Containers. *Journal of the American Medical Informatics Association* 2018, 25(1): 4-12. Source code available at <https://github.com/Bioconductor-notebooks>
5. In addition to facilitating reproducible bioinformatics analyses, my lab optimizes performance of big biomedical data workflows by leveraging algorithm design, implementation, cloud deployment and systems architecture of computationally intensive tasks.

- a. Zhang P., Hung L.H., Lloyd W., **Yeung K.Y.** Hot-starting software containers for bioinformatics analyses. *Gigascience* 2018, 7(8) giy092. Software available at <https://github.com/paizhang/Hotstarting-For-STAR-Alignment>
- b. Hung L.H., Lloyd W., Sridhar R.A., Ravishankar S.D.A., Xiong Y., Sobie E. and **Yeung K.Y.** Holistic optimization of RNA-seq workflow for multi-threaded environments. *Bioinformatics* 2019, btz169. Software available at https://github.com/BioDepot/LINCS_RNAseq_cpp
- c. Hung H.L., Kumanov D., Niu X., Lloyd W., **Yeung K.Y.** Rapid RNA sequencing data analysis using serverless computing. [bioRxiv 576199](https://doi.org/10.1101/576199) [Pre-print] 2019. Software available at <https://github.com/BioDepot/RNA-seq-lambda>
- d. Niu X., Kumanov D., Hung H.L., Lloyd W., **Yeung K.Y.** Leveraging serverless computing to improve performance for sequence comparison. Accepted for publication at the 8th *International Workshop on Parallel and Cloud-based Bioinformatics and Biomedicine (ParBio)* 2019.